Henry Greely
American Society of Law Medicine & Ethics
DNA Fingerprinting and Civil Liberties Project
RO1-HGH002836-02
Workshop 3 Presentation
May 3, 2005

My topic today is ethical and regulatory issues in DNA repositories and databases. Mark Rothstein, who is speaking shortly after me, is talking about further ethical issues in biobanks. Mark and I haven't actually had the opportunity to coordinate too well, but there shouldn't be too much overlap. I think Mark is going to focus more on some of the legal issues and I'm going to focus more heavily on some of the important ethical issues.

The topic that I want to talk about is not exactly DNA repositories and databases and it's not exactly forensic databases. It's broader than that. What we need to think about is general regulations for entities that encompass both genetic and other physical data, physical samples and health care information. I tried at one point to label these with the *very* catchy label of "genotype/phenotype resources" which, not surprisingly, went nowhere. Biobanks – as a sheer title, I don't think, is as informative. But it's these genotype/phenotype resources that we're interested in –created for forensic purposes, at least in part. I do think this is an area where the appropriate regulatory and ethical changes need to have a broad focus instead of just a narrow focus specifically on forensic issues.

I will speak to some forensic issues that are especially relevant but I want to talk more broadly about these entities that combine physical samples from individuals with information about them.

I have three initial points – and then I want to talk to you about 5 major issues.

The first one – it's really important – is that it's easy to get caught up thinking about the DNA, but the DNA has almost no value, medically and scientifically, unless you combine it with information about the subjects from whom the DNA is taken.

You might learn a little bit about population genetics broadly or the degree of variation in a particular gene, but if you're interested in any phenotype, in any trait, in any disease, in any behavior – you have to have information about the subjects that is connected to their DNA information. This health information or phenotypic information is often much more expensive, much more difficult, much harder to get and of much more dubious quality.

In some ways I think we make a mistake when we focus too much on the DNA part. The DNA part is really the tail of the dog. It's an important tail, and it's a dog that can't

be tailless, but the main part is really the phenotypic information.

My Second point is that there is a huge variety of things that fall into this broad sweep, ranging from the plans for the NHS database and repository for DNA and health information from 500,000 Britains to whatever decoding is finally done in Iceland, the individual labs that have 12 samples from individuals that they've done some analysis on. All of these are genotype/phenotype resources.

It also encompasses not just somatic sequence information, whether your inherited sequence or your mutated sequence in the case of many tumors, but there's this huge expanding area of RNA analysis where what's interesting is not so much the sequence but which parts of the sequence are being expressed at any time.

Databases and materials repositories that provide tissues for that kind of analysis are also a big and growing area that needs to be thought of here.

And then horizontally, as I think of it, you really have some strong similarities between these genotype/phenotype resources and all other sorts of medical research resources ranging from purely electronic health data information, to patient's clinical records done electronically and to research and pure samples given off with very little information.

All of those raise some of the same sets of issues and need to be dealt with in a unified way, or at least considered in a unified way.

Third and last of these preliminary points, and I think to me the most surprising is the analysis that Ellen Wright Clayton and her group published in 1995, throwing a grenade into the research world with their article in *JAMA* on the use of stored tissue samples. They pointed out that there were lots of legal and ethical problems lurking here to which people hadn't paid a lot of attention. Whether you agree entirely with her group's analysis or not, they raise lots of important issues. We're now almost to the 10$^{th}$ year anniversary of the publication of that paper and, at least in a legal sense in the United States, we're not very close to having settled legally any of the questions that were raised there or any of the ones that followed.

That is surprising because everybody knows this enterprise – and I use that term to encompass both non-profit and for profit parts of it – has expanded enormously in the last 10 years. If you ask a lawyer hard questions about any aspects of it, the honest answer will be "well, it's not clear. I'm not sure. I don't know. I can't be positive." Yet, the enterprise continues to expand despite the legal uncertainty.

At some point it would be nice to have legal certainty. On the other hand, since we've managed to expand for 10 years without disasters without this certainty, maybe things can keep going on in this awkward state forever. I would bet against it, but we'll have to see.

So five areas where I think the ethical issues in these genotype/phenotype resources are particularly pressing are issues of consent, issues of control, issues of confidentiality, issues of communication of results and issues of commercialization.

First – the issue of consent.

For the most part, the core question of individual consent to give up their samples for research hasn't been a big problem. You typically only get samples by new intervention which is required through regulatory mechanisms. One of the interesting things about the topic for this conference is that forensic databases don't involve consent, for the collection of the DNA itself, and neither do collections at mass disasters. One could imagine something like Craig Ventor's plan for trolling the seas and trolling the air for miscellaneous DNA. You could imagine walking out to Harvard Square and just collecting a bunch of miscellaneous human DNA. You wouldn't have any phenotype information with it so it wouldn't be particularly useful for most purposes. But you can do that without consent.

Then we get into the question of, since the collection is without consent, can the research be without consent? What kind of research is allowed with what kinds of consent? Those are particularly interesting questions in the forensics area and I'm sure that's a part of what we're going to be talking about.

With the decode experience in Iceland, they're not dealing with the collection of genetic samples. The Icelandic Parliament, the Alping, passed a law allowing the collection of health information with presumed consent, rather than affirmative consent. It was an opt out system. You were in unless you affirmatively sent a form to the government opting out. That kind of presumed consent did not apply legally to the collection of DNA samples in Iceland. As far as I know, the nationwide health database has never started up in Iceland, despite the fact that the license was issued over 5 years ago.

Cynics like to think that once decode went public at a nice stock price, their incentives to spend the money to build this database may have shrunk some. But that's a cynical lawyer's perspective.

So presumed consent is a model in some context, usually we deal with full individual consent, at least to have your DNA collected and be part of it. The issue that's more interesting in the straight consent area is one that I futilely argued about for almost a decade now, the question of group consent. When you're doing research with or building a repository of individuals from groups defined sociologically, culturally, ethically and any other particular way that, there is a role for group consent since the research results are likely to be about that group and not just about the individuals. Since the entire group could be benefited or harmed by the research results, whether or not the individual members chose to participate in that research, you should seek and require permission, both from the individual and from the group.

This argument floated by me for the Human Genome Diversity Project about nine years ago. It generally went nowhere, in part because of some ethical concerns about it, in large part because of the practical concerns and the practical difficulties, some of which, I think, are entirely real. As an Irish American I have no idea who would speak for Irish Americans. But if I wanted to do research with the Navajo Nation I have a pretty good idea who can speak, maybe not universally, but largely for the Navajo Nation, and that's the government of the Navajo Nation. With Old Order Amish communities, they have, internally understood if not legally recognized, authority structures that can be used.

Now this argument has gone on and on with respect to ethnicity. The NIH (National Institutes of Health) has adopted a compromised position of community consultation. It will be interesting to see whether this works or not. I have some fears that communities may not like it very much if the NIH requires people to consult with them to talk with them but not necessarily to listen to them and not to do what the community wants. So if you've got a community consultation where the community says "X", and the researchers go ahead do "not X", because they are not required to get consent, but just consult. I fear that that may do more harm than good for the research enterprise in general and in that particular case. But we will see. I think it's too early to see to tell yet how the community consultation model is going to work out. My Stanford colleague, Mildred Cho is actually involved in some empirical research on how community consultation actually is working.

One spin off of the group consent idea that hasn't gotten as much attention, but I think in some ways is even more interesting, focuses not on ethnicity, but on other ways in which groups are involved in research. One example might be diseases. If you are doing research on the genetics of Huntington's Disease, people with Huntington's Disease are all affected, or people at risk of Huntington's Disease, or people in families with Huntington's Disease are all affected by your research results whether they consented or not. And I have suggested that maybe we should talk to these organizations, talk to Parkinson's organizations, talk to the Autism organizations.

Consent is a harder concept there. I think they're more likely to do something more like consultation. Particularly in the case of the many diseases there are multiple organizations and each of them hates the others, consultation gets particularly difficult. However, the fact that it's difficult doesn't mean that we shouldn't try. The forensic context offers, I think, a particularly interesting viewpoint. If you're studying DNA collected from felons by a state, what group are you studying? Felons. Well, maybe you should talk to felons as a group, representatives of felons, focus groups with felons, surveys with felons, and figure out what they think about your proposed research as a group, as opposed to individually. I do believe that groups, when they get together and talk about things, think about things and have leaders come up, and have arguments about things will often end up with different results than the one-on-one interaction between subject and researcher.

Now, frankly, I suspect that felon groups wouldn't be all that keen on research with the

samples that they left. But I don't know. It's an empirical question. But it's one that I think is of particular interest here, if felons as groups aren't interested, it may raise some questions about the ethics of felons as individuals being participants in this kind of research.

Second Topic: Control.

In some ways, I think this blends in with consent, but I want to separate it out to focus on two particular aspects of control: control over what is done with the samples and control over who does it. This is not as academic a discussion as the consent discussion has turned out to be. There's currently litigation in Arizona, involving the Havasupai Tribe - one by members of the tribe as plaintiff against Arizona State University and a number of researchers at Arizona State.

One of the suits actually names Stanford University and one researcher at Stanford based on a claim that the researchers did research with them for one stated reason, to study diabetes, and used the research for other things without their knowledge or consent. According to the allegations of the complaint, there has been no trial; all we have is allegations of the complaint. I don't think there's been an answer yet. According to the allegations of the complaint, researchers affirmatively misled the tribe and lied to them about what they knew - they were doing schizophrenia research and lied to them about it. The complaint is what the plaintiff's lawyer hopes he will be able to prove, not what he actually knows to be true. If this proves to be true, it's very troubling. And it raises issues of consent, of control both over what's done and who does it.

The issue over what's done, I think is a pretty straightforward one. You give your DNA for a study on hypertension, and it gets used for a study on colon cancer. Well, you probably don't care all that much. It gets used on a study on schizophrenia and genetics, well maybe you care, particularly if you're a family with members with schizophrenia. It's a study on alcoholism. Maybe you care a lot, depending on your family. It's a study on race and I.Q. The control of subsequent uses of the DNA and the associated information that you give researchers is one in which, I think, reasonable research subjects can genuinely have understandable and appropriate concerns. Traditionally it hasn't been that big a deal because usually when the materials are collected for research on one thing, the phenotypic information is usually limited to that thing. If all of the information you got is about diabetes, you are not likely to be able to say anything very interesting about alcoholism or schizophrenia anyway. But the change in research towards larger genotype/phenotype resources changes this. The effort to get samples with information about everything to be used for everything makes it much harder to be confident about to what uses your materials will ultimately be put.

The example would be something like the National Health Services planned system in the UK, to collect and record 500,000 people, their broad health records, and their phenotypic information. You are not going to build that for the purpose of studying one disease, or even one organ system. Because the one thing we know for sure about

these resources is that they are going to be extraordinarily expensive, largely because of the collection of the phenotypic data. We are talking billions of dollars. We are not going to spend billions of dollars to create a resource that can only be used for one disease. You want to use it for everything. It makes sense economically. It's not economically rational to do much else. But then the problem is, how does somebody give informed consent in advance to a variety of uses, some of which have not yet been imagined, that might be made of their DNA and their health information sometime over the next century?

Now there are some significant legal issues here that I think Mark is going to touch on about HIPPA and the common rule. I only want to suggest, regardless of whether blanket consent is legal or not or under what circumstances blanket consent might be legal, I think that there are some other alternatives that deserve being explored. One of the easier ones to implement would be a checklist. Give the subject three pages of names of diseases and conditions and let them check off either the ones that they want to be tested for, that they are willing to have research done for, or an opt out system for ones for which they are not willing to consent. That's not deeply satisfying, though I think we all have some suspicion of just how much attention research subjects pay when they hit page 47 of the consent form.

Another one that a group I am involved with has implemented, I think perhaps has more promise, to talk generally about the kinds of diseases and conditions for which you intend to use the materials. However, to say that if somebody proposes research that a committee drawn from the subjects, or the researchers, that the committee believes the topic of this particular research proposal might be particularly sensitive to some of the research subjects, then and only then do you require recontact and reconsent.

There are two extreme positions. One is to allow blanket consent, giving researchers ability to do everything without ever needing reconsent. The other is to necessitate constant reconsent with each additional research project requiring new informed consent. Frankly, I think neither of those is attractive or feasible. Finding something in between, I think, is going to prove to be quite important ethically and then the greater difficulty is finding a middle ground legally, with our current laws.

The second control issue may seem minor, though it has been an issue in litigation: who gets the information and data. There is a long and generally quite useful tradition in science of sharing materials and information. You publish a paper, somebody else wants to replicate it, there is arguably an ethical obligation to give them the materials that they need to try to replicate it. To the extent that there has been tension on that understanding, it's been from intellectual property concerns and material transfer agreements which arguably have limited the useful sharing of information more than would be scientifically optimal.

But do research subjects ever know that? Does the research subject who trusts you and gives you her DNA, her medical information, her family health records and history,

know it may later end up in a different country, or with a different set of researchers who aren't the people she trusts? I think the answer to that is to make sure that people know that, but know it, not just in a way buried in the depths of page 63 of the consent form. People must really know that the information may well be shared. Some people will, I predict, choose not to participate if that's the case.

Both of these issues of control can be applied at the individual level as well as the group level. So that if one requires some sort of group permission, the group permission might also apply, not just to whether the research can be done, but on what topics and by whom. One of the causes of action in the Havasupai suit is that the materials were given to other researchers, the researcher at Stanford, without the knowledge or consent of the tribe or the individual research subjects.

Third issue: Confidentiality

There's a lot of really interesting discussion and work, legal and otherwise, on confidentiality issues. The National Bioethics Advisory Commission, which made a valiant stab at trying to resolve many of these legal issues, also took a stab at some of the confidentiality issues. There's widespread confusion about terminology, about what it means to be anonymous, anonymized, and to be linked and unlinked. It would be an enormous advance if we could agree on common terminology that everybody understands, but I don't know what that terminology should be, and I'm not going to try to do that right now.

There's another deeper confidentiality issue. I don't think it can be solved, but can be made part of the warning or alert to the subjects. Even if you use anonymous data, or data through a trusted system that assigns codes to samples to assure anonymity, you still can't promise confidentiality. At least not with any sophisticated, rich data set.

Latanya Sweeney at SUNY Buffalo has published some work on this. Latayna's work is too statistical for me and computer science oriented for me to feel that I understand it. But I think that these two examples work for me. Let's say you get an anonymous sample with no identifying information. All you know is the person's date of birth, place of birth, their entire health records that include the zip code they currently live in, their date of birth, their place of birth and all other information you'd expect. You get somebody whose zip code happens to be the White House, which has its own zip code, and whose age is exactly George W. Bush's age, and has a history of alcoholism in his twenties and thirties that stopped in his late thirties. I think he's also had some knee surgery. Even without an address or name or social security number, you can figure out that that's George W. Bush. But they don't have to be famous. Being famous just makes it easier. If you're somewhere in between, like most of us, it's still going to be relatively easy to identify you.

I was born on June 25, 1952 in Columbus, Ohio. I think a rich, health information data set will have that information about me because where I was born and when may turn

out to be medically significant. Last time I saw my ophthalmologist, he ask me, “Did you grow up in the Ohio River Valley?” because I’ve got histoplasmosis scarring on my retina, which if fairly common for people who grew up around the Ohio River.

The fact that I was born in Columbus may turn out to be relevant. The year I was born or the time of year I was born may turn out to be relevant. There are a variety of interesting pieces of research about seasonal effects and how they may affect health issues.

In Columbus, Ohio in 1952, there were probably six children born on June 25, given the population of city and the fertility rate at the height of the baby boom. Of those six, on average 3 were going to be female and the other 3 were going to be males. Of the males, probably 0.5 would have been black and 2.5 would have been white at that point in the history of Columbus.

If you went to the Franklin County records and looked, you could find the three males who were born on that day. And then it wouldn’t take very much more information to figure out which one of them was me, if you could track them subsequently. If the other two were of a different race on the birth certificate than I was, that would make it easy. I happen to know who one of the other ones is. He is my cousin Mike. He’s 5'8", 140 lbs; it wouldn’t take too much information to distinguish my cousin Mike from me.

So, for most of us right now, it’s not worth it to go through that effort. But with every passing month, more and more information gets put online, actually, a lot of it, because of genealogy, which is apparently the second biggest source of internet sites. Genealogists love to put public records online. More and more things are going online and each time it goes online, it becomes cheaper and easier to get this kind of information.

I think what that means is that you can’t guarantee complete anonymity, even if the research is completely anonymous. If it involves a rich database, you have to tell your subjects that there’s some chance that we won’t be able to protect your privacy. We won’t be able to protect your confidentiality. We’ll do the best we can, but we can’t be perfect - totally apart from whether they’re using trusted systems, anonymity, or any of the other buzz words that constantly get thrown around and confused.

Issue four: Communication of Results

One level of this is fairly easy and low key. I think to the extent you get scientific results or to the extent that you don’t get any scientific results, it’s at the very least polite to tell your subjects what you found or what you didn’t find. I think it helps build their support for research, makes them happy, and gives them some reinforcement that it was a good thing that they participated in research. But you can’t just mail them a copy of your article in *Nature and Genetics* and expect that to be comprehensible. You’ve got to provide it in a way that’s meaningful, that’s understandable at a lower level, so that

people can really understand what you found and what you didn't find. That may even mean translating it into the different languages. It may mean going through an intermediary from that community who can put in terms that the community understands better. There's a lot to communicate and are many tricky aspects to communication. My own sense is that scientists often don't try hard enough in part because they almost never have the money to allow them to try hard enough to really communicate their results.

But it's the second issue that I think is the bigger one, a land mine waiting to be stepped on in human genetics research. Let's take this hypothetical. You are doing research on a bunch of people. You've given them the normal informed consent, which almost always says, "We are not going to tell you any of the individual results of this research no matter what. Period. Don't ask, thank you very much." And it turns out that one of the subjects that you've got, you're sequencing a variety of genes and you find out he's got a mutation in his MSL1 gene, one of the ones responsible for heredity nonpolyposis colon cancer. This is a highly penetrant allele that is extremely likely to give you colon cancer. Colon cancer, of course is a nasty disease to have. But like a number of nasty diseases, the earlier you know you've got it, the better off you are. Here you have some prophylactic things you can do, not just monitoring. If you and your doctor agreed, you could even do a colectomy and take out the colon entirely. That's not the world's most pleasant procedure, but it's better than dying of metastatic colon cancer. There are interventions that may improve your health chances, but only if you know you are at risk. If the researcher doesn't tell this person that he is at risk, and subsequently he dies, and the family finds out that the researchers knew that he had this mutation and didn't tell him, I suggest that in the tort litigation and public relations fiasco that would follow, the defense that, "well, the informed consent form, which our lawyers advised us to draft, says we weren't going to return any information," isn't going to be very helpful. And ethically it shouldn't be very helpful. If you've got powerful information that can make a real difference in somebody's health status, you should feel an ethical obligation to return it.

That's very easily said. There are innumerable difficulties, ranging from legal issues over CLIA, the Clinical Laboratory Improvement Amendment Act of 1988, to real issues of how much information do you have to have, how penetrant is it? How well published, how well established does the correlation have to be? How hard do you have to look to find them? Research subjects have a way of becoming unfindable very quickly. Is one phone call to a listed number enough? Do you have to look in the phonebook? Do you have to hire a private detective agency? What's your risk of alarming people unnecessarily? All of these are tough questions, but they are all questions of implementation of what I think is an ethical obligation that really can't be shirked.

A researcher in particular should have a duty to warn and a duty to prevent illness when he clearly can, or at least to mitigate bad health effects when he clearly can. He's not a disinterested third party. And to the extent that we continue our current legal and general practice of not returning information, even though there are lots of good

practical reasons for it, we run the risk of having an ethical fiasco, a legal fiasco, and a public relations disaster for genetics research.

Last Point: Commercialization

The commercialization issue is actually one of the hardest here. There is an argument, and I think an argument that has a lot of force, that complaints about commercialization are usually the outgrowth of other problems, just as most malpractice suits don't arise just because of a bad result, but because of a bad result plus a bad doctor/patient relationship. Commercialization concerns reach the level of litigation usually when it's not just a concern about being cheated out of money, but a concern about being cheated, lied to, and mistreated in other ways. Maybe we should really focus on the "cheated, lied to, mistreated in other ways" and not worry so much about commercialization. And I think that there's a plausible argument for that, assuming one could successfully eliminate the other forms of mistreatment. Still, though, I think at the core there is a specific issue about commercialization involving some of the ways in which the research world has changed.

In the golden age, which of course never existed, but only exists in memory, researchers were altruistic, weren't concerned with money or stock options, and research subjects were altruistic, only wanted to help humanity, weren't concerned about finding cures for their families or anything else.

Of course that golden age never existed. The opposite, the age of dross, in which some people paid us today also doesn't exist, even researchers with stock options also want to help people and want to advance knowledge. Besides, many researchers don't have stock options.

But it is the case that biomedical research has become much more lucrative for researchers. That's not necessarily a bad thing, but it does mean that the research subject signing the consent form deciding whether to participate in research or not, could honestly say, 'You know, you stand to make millions of dollars if you find anything out of this. What's in it for me? You can't do it without my material, my participation, or at least that from people like me. If you're going to profit financially, then why shouldn't I?"

Now I actually think that it's not so much an ethical obligation to share some of the profits, as it is a desire to avoid having people feel that they're suckers. You know, "I gave away something for free that you got awfully rich from. You took me for a ride. You mistreated me. I was a sucker" - which, whether or not it's a strong ethical argument, doesn't have good political consequences for support of research, I think.

Now where we go from there is complicated and difficult, Barth and the HUGO Ethics Committee proposed several years ago, really a pioneering effort that companies have an obligation to return some share of their revenues to good deeds, more generally with

places where they've done research. Others have argued that you should try to do good things for the communities in which you've done the research, sponsor a clinic, build a park, do something like that. Others have argued that individuals should be able to cut their own deals and take money or royalty shares.

The latter runs into some sort of a flip side problem. People are all too willing to believe that there are bonanzas circulating through their veins. The actual estimated present value of any one person's DNA and information in any one research project is probably well under a penny. To the extent that you are selling them, giving them a royalty and selling them a lottery ticket, there's the possibility that it will prove an undue inducement, that people will think that they are getting something of much more value than they actually are. Then they'll be annoyed and embittered when, as it will turn out to be the case 99.9 times out of a 100, there's no payoff.

So finding how to return a benefit is, I think, a difficult issue. Giving gifts is always hard, as I remember each Christmas, Valentines Day, anniversary and my wife's birthday. It's harder for some. It's harder for adults than for kids. Returning benefits and figuring out what benefits to give can also be very difficult, but I think it's something that researchers should feel ethically compelled to do. You can't do your research without this contribution. It may not be as important a contribution as your brains, or your ABI sequencing machines, or your data bases, but it's a necessary contribution without - DNA, or even more importantly, health information, family histories and so on, you just can't do it. And if there's going to be any financial return, it seems to me only fair that the people who provide this essential part of it not feel cheated, if they see you get a huge financial return while they get nothing.

And so I'll conclude by saying that all of this, for me, really adds up to this: I think that it is very important for research to avoid creating embittered research subjects. Having research subjects who fairly or unfairly, reasonably or unreasonably feel that you've betrayed them, you've lied to them, and cheated them raises a red flag about whether you've acted ethically. And even if you have acted perfectly ethically, and your conscience is clean, it's not a good thing politically for science to have unhappy research subjects out there. They're not going to participate in research again. They're not going to let their friends and relatives participate in research any time soon. And they are not going to be particularly interested in supporting candidates who say, "Well what we really need is more biomedical research."

I think both for ethical reasons and pragmatic reasons, it is important for us to get these rules right. Genotype/phenotype resources, whether they're in the forensic context or otherwise, don't raise unique issues here, but I think they take the issues that have always been there and bring them into particularly sharp focus.

So my own view is we need to work toward a sustained ethical and legal framework to make sure that this kind of research is done in a way that is both fair to the research subjects and perceived by them as being fair to research subjects.

Thank you.